基于矩阵分解学习的科学合作网络社区发现研究*

施晓华 [1,2]　卢宏涛 [2]
[1](上海交通大学图书馆　上海 200240)
[2](上海交通大学计算机系　上海 200240)

摘要:【目的】在科学合作网络的发展及主要社区发现方法的基础上，提出发现合作网络社区信息的方法。【方法】以情报领域部分相关期刊 2012 年–2016 年发表论文的共著网络为实验数据，基于贝叶斯对称非负矩阵分解方法，结合自动相关确定稀疏压缩原理，实现社区数量的自动获取，并在分解过程中应用对称矩阵分解原理。【结果】通过与现有方法的比较与分析，本文方法得到较好的实验结果。【局限】网络数据获取中未引入学者甄别的优化方法。【结论】本文提出的方法能有效解决合作网络社区发现需求。

关键词: 科学网络　文献共著网络　社区发现　非负矩阵分解　贝叶斯方法
分类号: G252

1　引　言

社区发现是进行社会网络分析的重要方法之一，通过对大型复杂网络内部的社区发现，可以利用已有信息了解网络的运行情况，有效理解和分析整体网络的属性，进而发现最有关联性的网络内部组织来简化全局结构，理解网络的拓扑属性。网络中的社区可以通过图分割法[1]不断最小化社区之间的连接数获取；也可以通过如基于某个目标函数(如模块度)随机优化[2]及统计推理[3]等方法来实现。

随着知识和科学技术不断全球化发展，科学研究涉及学科领域不断增多，使得合作研究成为科学研究的主流方式。在科学网络(Scientific Network)中[4-6]，科学家或研究学者成为关键节点，他们通过各种不同的关联方式形成以科学协作为主的相互联系，如共同发表论文、发表文献的相互引用、在相同领域学科杂志或会议上发表成果等。通过将科学网络进行有效社区发现和聚类，进而简化整个庞大、复杂的科学知识系统；可以将作者聚类至不同研究社区中，社区内的人员有相同的研究方向或兴趣(强关联性)，不同社区之间的人员有不同的研究兴趣(弱关联性)，进而通过对不同社区的特性进行分析，获取不同作者所在的科学知识社区，理解科研人员的合作和交流模式，挖掘科研人员的研究兴趣；并可进一步应用图分析和可视化功能，从网络分析角度对合作数据有更加深入的揭示。

本文将提出一种基于矩阵分解学习的优化社区发现算法，从整体网络中发现内在的多个社区，并在实际网络中取得了良好的社区发现效果；同时通过基于贝叶斯推理的稀疏压缩方法，自动获取社区的数量，解决了一般矩阵分解学习中无法自动确定分解维度的问题。

2　科学网络

科学网络研究主要研究知识体系中科学家之间的社区特性，分析其相互关联性和演化趋势，常见的科学网络有出版物引文网络[7]、期刊共引[8]和耦合网络[9]、共词网络[10]和共著网络[11]。在共著网络中，学者成为

通讯作者：施晓华，ORCID: 0000-0002-8297-0750, E-mail: xhshi@sjtu.edu.cn。
*本文系上海交通大学 2013 年文理交叉项目“科学网络中知识社区发现技术与应用研究”(项目编号：13JCY14)的研究成果之一。

网络中的每一个节点，学者 *u* 和 *v* 共同完成的科学成果成为之间的链接(如论文、图书、报告及基金项目等)。对应的链接权重 $w(u,v)$等于 *u* 和 *v* 共同贡献的作品数量。在合作网络中，一般认为 *u* 和 *v* 对于作品的贡献是相同的，因此该网络是一个无向图，即 $w(u, v)=w(v, u)$。

Newman 在 2004 年分析了生物、物理和数学三个不同学科的论文数据，分析共同发表论文的作者关系(Coauthorship)，分析作者之间协作关系的社区结构，以及不同学科科学协作网络的差异，同年他还提出基于模块度的层次社区结构分类方法[12]；王福生等针对《图书情报工作》杂志 2001 年–2006 年的论文数据，构建，并分析作者科研合作网络模型的分布与演化过程[13]；Mimno 通过建立文献社区–作者–主题模型(Community-Author-Topic)提取科学家合作网络中的研究社区，并对 NIPS 会议上发表的论文数据进行科学社区发现的实验分析[14]；Erfanmanesh 等通过对 *Scientometrics* 期刊 1980 年–2012 年的 3 125 篇文章进行共著作者分析研究，发现其中的活跃、中心和协同节点[15]。

3 社区发现方法

随着复杂网络学科的不断发展，众多学者提出用以解释各种物理现象的网络模型，例如小世界模型、无标度网络以及随机网络等；同时伴随海量数据分析与挖掘为宗旨的数据挖掘技术研究的深入，计算机科学家根据信息网络的规模巨大性等特点，设计出高效率并具有一定智能的鲁棒社区分类算法，这些方法主要为基于图论的社区发现算法，包括谱聚类分割、层级社区发现算法和基于随机行走的聚类算法等。目前常用的网络处理工具有 UCINET、iGraph、Gephi、Pajek 和 CytoScape 等，不仅可以实现网络中节点和链接的定量分析，并支持简单的社区发现功能。网络社区发现根据聚合原理主要有以下方法[16]:

(1) 图分割法。网络的社区发现就是将网络按其内在的社团结构划分成一个个子网络的过程。在计算机科学领域，这类问题一般称作图分割(Graph Partitioning)。基于图分割的社区发现算法利用节点间关系的信息，其隐含的基本假设是：子社区内部的交互要远远比子社区之间的交互更为紧密。开源工具的 CFinder 和 NetworkX 使用 CPM(Clique Percolation Method)[17]完全子图过滤方法实现社区发现，其他的图分割方法主要还有层次聚类如 BGLL 算法[18]和矩阵分解学习方法等。

(2) 分裂方法。网络的社团结构分析往往要求算法能够准确、自动地确定网络的社团个数并给出相应社团的“自然”分割。近年来有大量的更适合网络社团分析的有效算法被提出。Newman 等提出一种通过边移除按层次地分解网络的社团分析方法[2]，这项工作被认为是现代社区结构分析方法的开创性工作，引起了各领域研究人员的广泛兴趣。

(3) 合并方法。在合并方法中，模块度(Modularity)[19]标准被提出并用于衡量社团结构划分好坏。网络的某种划分对应模块度值越高往往表明该划分越可能是符合网络社团结构的划分，一般社区的模块度都在 0.3 至 1 之间。模块度是基于随机网络不存在社团结构的假定，以所有节点的度值与给定网络相同但边随机连接的网络为参考模型，比较给定网络的所有社团内部边的数量与相对应的社团内部边的期望数量。Le Martelot 等提出基于模块度优化的快速迭代社区发现方法[20]，通过对已有社区不断增加节点，发现更高模块度的新社区分配。

尽管在理论和实践上基于图分割的社区方法取得了一定的成功，但仍然有一些不足。这些算法通常等价于一个特征值分解问题，而社区可以从生成特征向量确定，然而这些关键的特征向量并没有具体的物理含义。通常来讲，所有的网络结构均可以通过关系图来表示，其主要结构表征即为其邻接矩阵(Adjacency Matrix)，由于邻接矩阵的非负特性(所有元素为 1 或 0)，因此在机器学习领域研究的热点非负矩阵分解(Non-negative Matrix Factorization, NMF)得到了很好的应用。NMF 是一种比较新的矩阵分解算法，主要用于数据聚类应用，于 1999 年才被开始提出[21]，它克服了传统矩阵分解的很多问题，通过寻找上下文有意义的解决方法，提供解释数据的更深入看法，为人类处理大规模数据提供一种新的途径。因此将 NMF 应用到社区发现中，在分解过程中保持网络的非负性，往往能达到表示社区的局部之间相关关系的效果；并从全局出发，快速发现不同社区的信息，取得更好的社区分类效果[21]。

4 非负矩阵分解学习在科学网络社区发现中的应用

NMF 的基本思想可以简单描述为: 对于任意给定的一个非负矩阵 X, NMF 算法能够寻找到一个非负矩阵 U 和一个非负矩阵 V, 使得满足 X 和 UV 之间的最优近似, 从而将一个非负的矩阵分解为左右两个非负矩阵的乘积, $X \approx UV^T$ 。NMF 初始被应用于图像分类、文本挖掘等领域, 近年来有不少研究者开始将这种方法应用于各种网络的社区发现中[22], 其中包括大量的科学网络数据。Wang 等[23]通过对第 0-12 卷的 NIPS 会议论文数据进行合著网络分析, 有效发现社区中的重要作者节点(Superstars)。Zhang 等[24]使用贝叶斯 NMF 三分解对 Newman 等的网络科学合作网络和高能理论合作网络进行社区发现实验。Mankad 等[25]使用加入 L1 范数稀疏约束的 NMF 方法对 2003 KDD Cup 提供的引文网络数据进行社区发现。Yang 等[26]通过 NMF 方法从大规模 DBLP(计算机科学协作网络)网络中有效发现其中的重叠社区结构。

Xie 等[27]通过调研社区发现方法, 提出 NMF 处理社区发现问题时, 其分解维度设置和复杂度简化是主要问题:

(1) NMF 是一种无监督的方法, 通常社区的数量 K 是未知的, 因此需要有效判断社区数量 K;

(2) 对于一些特殊网络, 如合著网络, 源图是一个无向图, 需要充分利用网络结构优化分解算法, 提升算法效率。

如图 1 所示, 本文将以自动收割到的期刊发文与作者数据为基础, 生成合著网络(对称网络), 并以邻接矩阵形式表示, 并通过贝叶斯对称非负矩阵分解(Bayesian Symmetric NMF, BSNMF)[28-29]进行实际的社区发现实验。

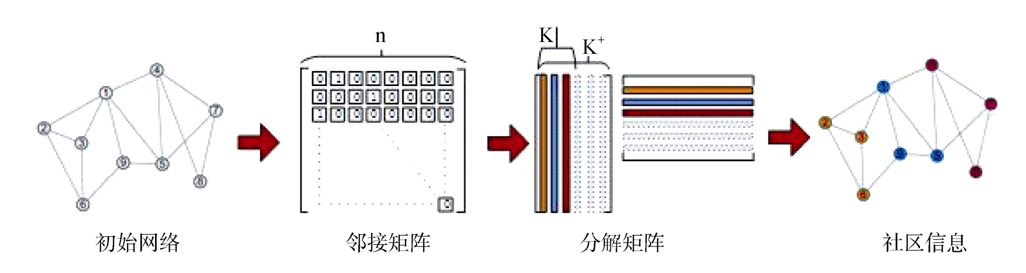


图 1 BSNMF 过程演示

(1) 通过贝叶斯假设和 Automatic Relevance Determination(ARD)稀疏方法[30], 有效判断获取社区的数量 K。ARD 是一种基于贝叶斯推断的模型选择方法, 它通过超参数的应用, 表示模型对应不同特性的相关性, 并通过定义一些参数来描述这些特性的偏差范围。如果偏差为零, 则相应的特性对模型预测不起任何作用。因此, ARD 可以自动发现模型中相关的一些关键元素, 在社区发现中, 即为分解出来的主要社区属性。

(2) 设定分解的两个矩阵是对称的, 既 $U = V^T$, 大大简化计算过程。

BSNMF 的算法过程如下:

输入网络邻接矩阵 $X \in R_+^{N\times N}$, 初始化 K, 固定超参数 a,b;

定义矩阵 $B \in R_+^{K\times K}$ 是一个对角线为 β_k 的对角矩阵;

① 初始化矩阵 S_0

② 从 i=1 到 n_{iter}

③ $S \leftarrow \left(\frac{S}{1\times S + S\times B}\right)\cdot\left[\left(\frac{X}{S\times S^T}\right)\times S^T\right]$

④ $\beta_k \leftarrow \frac{N+a-1}{\sum_i s_{ij}^2 + b}$

⑤ 结束

⑥ $K_* \leftarrow$ S 中非零的列数

⑦ $S_* \leftarrow$ 去除 S 中全零列

⑧ 返回 K_*和 S_*

其中, X 为输入的网络邻接矩阵, K 为初始设置的社区数, S 为分解后的目标矩阵, S_*和 K_*是对 S 去除全零向量后的结果; a,b 是模型的超参数, 通过多次实验比较确认, n_{iter} 为迭代次数。

通过抽取, 建立网络矩阵 X, 在 MATLAB 下运行 BSNMF 算法, 不仅自动获取到社区的数目 K_*, 对于返回的 S_*的维度为 $N\times K_*$, 取 S_*中每一列的最大值 $c=\mathrm{argmax}_c S_{*ic}$, 即为第 i 个点所属的社区信息, 自动获

取每个社区的信息。

4.1 数据处理

通过中国期刊网的 RSS 聚合源，获取《情报资料工作》、《情报理论与实践》、《情报科学》和《情报杂志》4 种情报领域期刊的 2012 年–2016 年的论文和作者信息[31]，共计 6 254 篇论文，对应作者 14 312 个。再通过网页抓取方法，获取论文对应的单位信息，以作者姓名结合作者单位作为去重条件，去重后得到 7 625 个作者。每两个作者有合作发表论文，则两点之间有边相连，边权重为两个作者合作的篇数；据此形成一个 7 625 个节点、12 672 条边的大规模科学合作网络，其中每个节点平均度为 1.66，即 5 年中每人平均和 1.66 人次同行进行合作发文。

表 1 显示了发表论文最多的 10 位作者在合作网络中的度数信息。由于不同人员合作发表论文的情况不同，若一位作者全部以独立作者进行发文时，其在合著网络中表示为一个独立节点，经计算，在此网络中共有 665 个孤立节点，因此在进行社区发现应用之前，先将这些孤立节点删除，以提升整体社区发现的效果。形成新的合作网络 X 包含 6 960 个节点，边数仍然为 12 672 条。

表 1 发表论文前 10 的作者信息

姓名	单位	发表次数	网络度数
邱均平	武汉大学	50	66
朱庆华	南京大学	41	83
黄鲁成	北京工业大学	40	118
赵蓉英	武汉大学	36	49
陈福集	福州大学	35	48
王国华	华中科技大学	30	89
谢阳群	淮北师范大学	27	48
娄策群	华中师范大学	26	44
张玉峰	武汉大学	26	37
孙建军	南京大学	25	42

4.2 主要方法比较

结合生成的合作网络，将现有的多个社区发现方法进行模块度结果比较。

(1) CPM 完全子图过滤方法[16]，使用 3-Clique 发现网络中的完全子图数量，未发现的节点作为独立节点来计算整体模块度；

(2) GN 算法[18]，基于聚类中的分裂原理，使用边介数作为相似度的度量方法；

(3) BGLL 层次化社团结构的凝聚算法[17]；

(4) Louvain 贪婪优化方法[19]，在迭代中不断优化分类社区的模块度；

(5) NMF[20]，利用非负矩阵方法的聚类功能实现社区发现；

(6) SNMF[22]，对称非负矩阵方法处理无向网络的社区发现；

(7) BSNMF，本文提出的贝叶斯对称非负矩阵分解方法。

BSNMF 方法设置初始社区数为节点总数的五分之一，初始设置超参数 a=5, b=2，运行 n_{iter}=500 次迭代以保证收敛。在获得矩阵 S 中，对应的非零列共有 702 列，即共分成 702 个社区。由于方法 2、方法 5、方法 6 需要初始设置社区数，实验设定为由 BSNMF 方法获取的社区数 702。

经验证，在表 2 的已有算法中，Louvain 算法取得了比较好的社区发现效果，而 BSNMF 方法发现的社区结果模块度达到 0.9664，得到最高的网络分区效果，同时，相比方法 5 和方法 6，BSNMF 中的对称矩阵分解方法和贝叶斯方法都取得了更好的实验结果。

表 2 社区发现结果模块度比较

方法	模块度
3-Clique	0.3579
GN	0.5530
BGLL	0.8294
Louvain	0.9165
NMF	0.4209
SNMF	0.8165
BSNMF	0.9664

4.3 社区发现结果分析

图 2 为 BSNMF 发现的 702 个社区的节点直方图，其中紧密的前 6 个大型社区在表 3 中显示，在这些社区中的节点人员通过共同发表论文关系，形成不同紧密的作者合著社区或合著小组，在这些社区中的作者具有类似的研究兴趣或方向，今后这些作者之间有更大的可能性来继续合作发表论文。

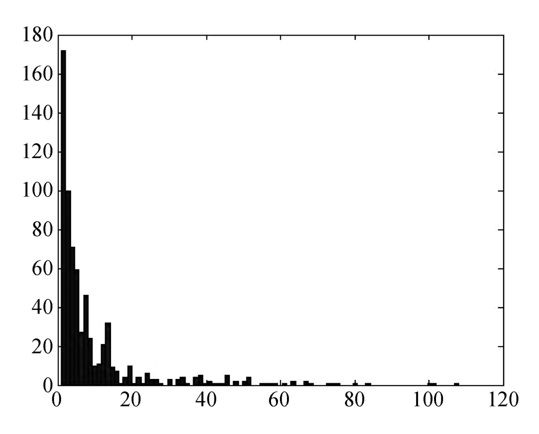


图 2　利用 BSNMF 发现的 702 个社区的节点数统计

图 3 通过网格图显示了社区 1 的主要节点作者关系图，形成了以孙建军、俞立平、郑彦宁等为中心节点的紧密合作社区。通过分析发现，从社区 1 共同发表的 140 篇论文中抽取得到的关键主题词有"情报"(34 次)、"评价"(26 次)、"网络"(22 次)以及"信息"(21 次)。

表 3　其中发现的几个主要社区及节点信息

社区	节点数	节点度和	主要节点人员(度大于 10，下划线为大于 20)
1	103	200	孙建军，俞立平，郑彦宁，潘云涛，武夷山，丁堃，姜春林，刘志辉
2	100	192	朱庆华，袁勤俭，宗乾进，赵宇翔，刘璇
3	100	220	黄鲁成，翟东升，苗红，吴菲菲，张杰，娄岩
4	89	159	毕强，彭洁，滕广青，黄微
5	84	194	王国华，曾润喜，钟声扬，陈强，王雅蕾，杨腾飞，徐晓林，张韦，闵晨
6	84	177	张海涛，徐宝祥，张连峰，崔金栋，武慧娟，王欣，王丹，许孝君，宋拓

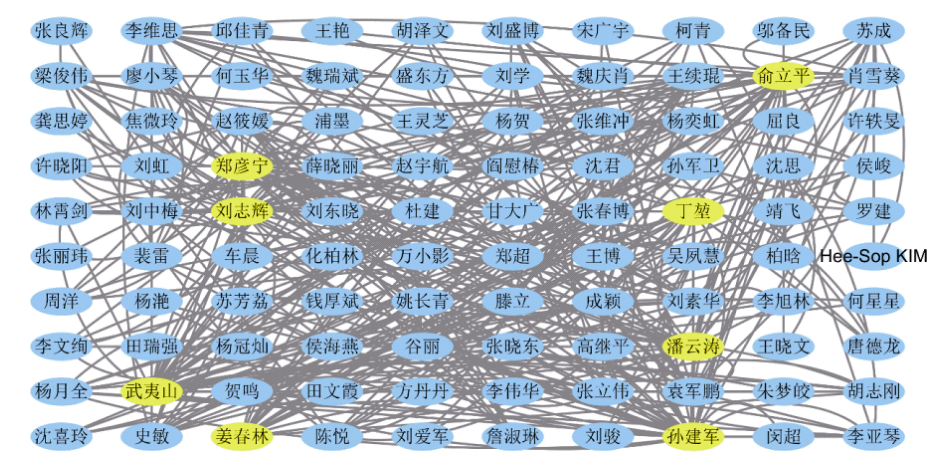


图 3　以孙建军等为主要节点的网格形社区结构(黄色节点为度大于 10 的节点)

如图 4 形成的网络合作发文主题词云图所示，在社区 1 中的人员更加关注情报学相关的网络评价和信息管理方法领域，同时人员之间也有更高的合作发文趋势。

图 5 通过环形图显示了社区 3 的主要节点作者关系图，形成以黄鲁成为中心节点的紧密合作社区。通过分析社区 3 共同发表的 116 篇论文，其中 69 篇集中发表在同一类期刊上，抽取到的关键主题词为"专利"(38 次)，说明社区 3 中主要的研究领域是专利分析为中心的相关文献计量和技术应用研究。

图 3 的社区 1 分布结构与图 5 社区 3 有所不同，社区 3 基于节点"黄鲁成"分成两个不关联社区，分别是以"翟东升"为共同主要节点和以"苗红，吴菲菲"为共

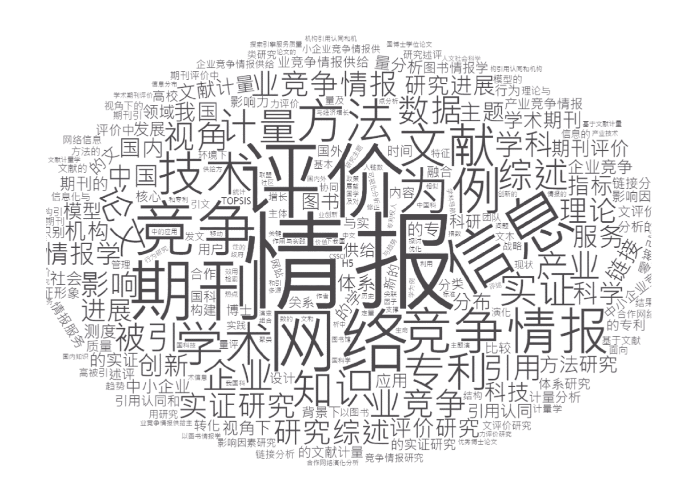


图 4　以孙建军等为主要节点合作论文的主题词云图

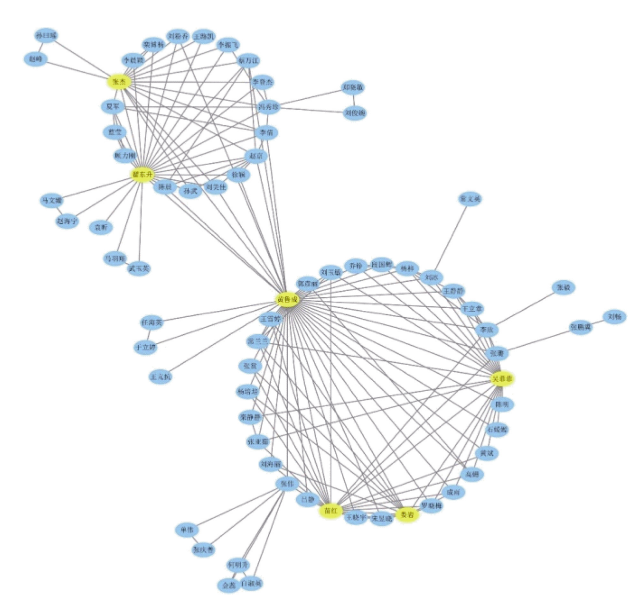


图 5　以黄鲁成为主要节点的环形社区结构
(黄色节点为度大于 10 的节点)

同主要节点的两个独立社区，因此后续可以使用重叠社区分区方法，将“黄鲁成”同时分布在两个不同社区，以达到更好的社区分类效果。

5　结　语

科学网络社区发现可以在大量文献中，有效发现潜在的学者关联社区，对基于网络科学的文献计量有很好的技术支撑作用，应用贝叶斯非负矩阵方法能快速有效发现相关社区信息，自动确定社区数量，并以情报学领域合作网络中进行应用实验，对比其他的社区发现方法，取得了更高的模块度值。同时一些节点可能同时出现在不同社区(见图 5)，出现重叠现象(Overlapping)，如何能自动、有效发现重叠社区，可以进一步拓展应用。在合作网络分析中，如何自动有效甄别作者身份信息[32]，形成和自然人的一一对应，是合作网络的数据清洗的关键，也是实现在跨学科、交叉领域学者社区发现的关键，需要在数据处理过程加入更多专家经验和智能化手段以取得新的突破。

参考文献：

[1] Ding Y. Community Detection: Topological vs. Topical[J]. Journal of Informetrics, 2011, 5(4): 498-514.

[2] Newman M E J, Girvan M. Finding and Evaluating Community Structure in Networks[J]. Physical Review E, 2004, 69: Article No. 026113.

[3] Tang X, Xu T, Feng X. et al. Uncovering Community Structures with Initialized Bayesian Nonnegative Matrix Factorization[J]. PLoS ONE, 2014, 9(12): Article No. e107884.

[4] 徐玲，胡海波，汪小帆. 一个中国科学家合作网的实证分析[J]. 复杂系统与复杂性科学, 2009, 6(1): 20-28. (Xu Ling, Hu Haibo, Wang Xiaofan. Empirical Analysis of a China Scientists Collaboration Network[J], Complex System and Complexity Science, 2009, 6(1): 20-28.)

[5] Bailón-Moreno R, Jurado-Alameda E, Ruiz-Baños R. The Scientific Network of Surfactants: Structural Analysis[J]. Journal of the American Society for Information Science and Technology, 2006, 57(7): 949-960.

[6] Quattrociocchi W, Amblard F, Galeota E. Selection in Scientific Networks[J]. Social Network Analysis and Mining, 2012, 2 (3): 229-237.

[7] Havemann F, Scharnhorst A. Bibliometric Networks[OL]. arXiv PrePrint, arXiv: 1212.5211.

[8] Chen P, Redner S. Community Structure of the Physical Review Citation Network[J]. Journal of Informetrics, 2010, 4(3): 278-290.

[9] 马瑞敏，倪超群. 作者耦合分析：一种新学科知识结构发现方法的探索性研究[J]. 中国图书馆学报，2012，38(2): 4-11. (Ma Ruimin, Ni Chaoqun. Author Coupling Analysis: An Exploratory Study on a New Approach to Discover Intellectual Structure of a Discipline[J]. Journal of Library Science in China, 2012, 38(2): 4-11.)

[10] 张斌. 共词网络的结构与演化：概念与理论进展[J]. 情报杂志，2014，33(7): 103-109. (Zhang Bin. Structure and Evolution of Co-word Network: Concept and Research Review[J]. Journal of Intelligence, 2014, 33(7): 103-109.)

[11] 苗蕊，刘鲁. 科学家合作网络中的社区发现[J]. 情报学报，2011，30(12): 1312-1318. (Miao Rui, Liu Lu. Community Detection in Scientific Collabration Network[J]. Journal of the China Society for Science and Technical Information, 2011, 30(12): 1312-1318.)

[12] Newman M E J. Coauthorship Networks and Patterns of Scientific Collaboration[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(S1): 5200-5205.

[13] 王福生，杨洪勇. 作者科研合作网络模型与实证研究[J]. 图书情报工作, 2007, 51(10): 68-71. (Wang Fusheng, Yang Hongyong. Author Collaboration Network Model and Demonstration Study[J]. Library and Information Service, 2007, 51(10): 68-71.)

[14] Mimno D. Community-based Link Prediction with Text[C]//

Proceedings of the 21st Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada. 2007.

[15] Erfanmanesh M, Rohani V A, Abrizah A. Co-authorship Network of Scientometrics Research Collaboration[J]. Malaysian Journal of Library & Information Science, 2012, 17 (3): 73-93.

[16] Fortunato S. Community Detection in Graphs[J]. Physics Reports, 2010, 486(3): 75-174.

[17] Palla G, Derényi I, Farkas I, et al. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society[J]. Nature, 2005, 435(7043): 814-818.

[18] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): P10008.

[19] Newman M E J. Modularity and Community Structure in Networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23): 8577-8582.

[20] Le Martelot E, Hankin C. Fast Multi-scale Detection of Relevant Communities in Large-scale Networks [J]. The Computer Journal, 2013. DOI: 10.1093/comjnl/bxt002.

[21] Lee D D, Seung H S. Learning the Parts of Objects by Non-negative Matrix Factorization [J]. Nature, 1999, 401(6755): 788-791.

[22] 李亚芳，贾彩燕，于剑. 应用非负矩阵分解模型的社区发现方法综述[J]. 计算机科学与探索，2016，10(1)：1-13. (Li Yafang, Jia Caiyan, Yu Jian. Survey on Community Detection Algorithms Using Nonnegative Matrix Factorization Model [J]. Journal of Frontiers of Computer Science and Technology, 2016, 10(1): 1-13.)

[23] Wang F, Li T, Wang X, et al. Community Discovery Using Nonnegative Matrix Factorization[J]. Data Mining and Knowledge Discovery, 2011, 22(3): 493-521.

[24] Zhang Y, Yeung D Y. Overlapping Community Detection via Bounded Nonnegative Matrix Tri-factorization [C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 606-614.

[25] Mankad S, Michailidis G. Structural and Functional Discovery in Dynamic Networks with Non-negative Matrix Factorization[J]. Physical Review E, 2013, 88(4): 042812.

[26] Yang J, Leskovec J. Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach[C]// Proceedings of the 6th ACM International Conference on Web Search and Data Mining. ACM, 2013: 587-596.

[27] Xie J, Kelley S, Szymanski B K. Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study[J]. ACM Computing Surveys, 2013, 45(4): 43.

[28] Psorakis I, Roberts S, Ebden M, et al. Overlapping Community Detection Using Bayesian Non-negative Matrix Factorization[J]. Physical Review E, 2011, 83(6): 066114.)

[29] Shi X, Lu H. Community Inference with Bayesian Non-negative Matrix Factorization[A]// Web Technologies and Applications[M]. Springer International Publishing, 2016: 208-219.

[30] Mørup M, Hansen L K. Automatic Relevance Determination for Multi-way Models[J]. Journal of Chemometrics, 2009, 23(7-8): 352-363.

[31] RSS-CNKI [EB/OL]. [2017-03-06]. http: //rss.cnki.net/rss.

[32] Tang J, Fong A C M, Wang B, et al. A Unified Probabilistic Framework for Name Disambiguation in Digital Library[J]. IEEE Transaction on Knowledge and Data Engineering, 2012, 24(6): 975-987.

作者贡献声明:

施晓华：采集、清洗和分析数据，完成实验，撰写论文；
卢宏涛：提出研究思路，修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: xhshi@sjtu.edu.cn。

[1] 施晓华. Raw_data.xlsx. 合作社区原始数据.

[2] 施晓华. Adjacent_Matrix.mat. 社区邻接矩阵数据.

收稿日期: 2017-04-10
收修改稿日期: 2017-07-04

Detecting Community in Scientific Collaboration Network with Bayesian Symmetric NMF

Shi Xiaohua[1,2] Lu Hongtao[2]
[1](Library of Shanghai Jiaotong University, Shanghai 200240, China)
[2](Computer Science Department, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: [**Objective**] This study proposes and examines a new method to identify the communities in collaboration network of scientific researchers. [**Methods**] First, we retrieved the need data from information science journal articles published from 2012 to 2016. Then, we used the Automatic Relevance Determination to find the target community with the Bayesian Symmetric Non-negative Matrix Factorization method. Finally, we compared the performance of our method with the existing ones. [**Results**] The proposed method got better results than others. [**Limitations**] Did not optimize our data with the researcher identifications. [**Conclusions**] The proposed method could effectively find communities from the scientific collaboration network.

Keywords: Scientific Network Co-author Network Community Detection Non-negative Matrix Factorization Bayesian Approach

SSRN 发布新的生物网络工具——BioRN

科学健康领域的全球信息分析公司——爱思唯尔宣布世界领先的工作底稿知识库和预印本服务器 SSRN 发布生物研究网络——BioRN。

生物研究者可以将预印本和工作底稿发布到 BioRN 上，分享观点和其他早期研究成果并展开合作。BioRN 让用户迅速上传以及免费阅读摘要和论文全文。

SSRN 总经理 Gregg Gordon 说: “经过几周的测试，对 BioRN 的反馈不同凡响。SSRN 处于一个特殊位置来服务生物研究社区。在分享方面我们已经有超过 20 年的经验，同时我们也从爱思唯尔的知识、技术以及与作者的联系中获益。无论你是一位生物学专业的学生还是高级研究者，在 BioRN 上面分享你的工作和合作将有很多优势。”

建立这个致力于促进生物发展的网络符合 SSRN 的发展逻辑，因为它拥有跨越诸如生物技术、生物伦理法律和生态等领域各个网络的大量的生物学方面的内容。BioRN 已经拥有来自 6 500 位作者的近 4 500 篇文章。这个全新的网络将通过促进生命科学与社会科学之间跨学科的合作、鼓励同行建设性的反馈、提升研究者的声誉来发展生物学研究社区。

SSRN 在其 eLibrary 拥有由超过 340 000 作者提交的近 625 000 份论文和 750 000 份摘要，有超过 1 亿的下载量。它为全球的研究者、学生、创业者、学者、企业和组织提供服务。当研究成果尚未在期刊或书籍上发表出来之前，SSRN 就为这些成果提供了分享和传播的机会。超过 200 万研究者使用 SSRN，并参与 30 个社区驱动的网络，涉及的学科包括经济学、管理学和法学。

一份预印本是作者自己对研究结果和尚未经过同行评议或接受任何由出版商提供的附加值(例如调整格式、副本编辑、技术增强)的分析的重新整合。一个预印本服务器或者工作底稿知识库让用户可以分享这些文件。

(编译自: http://www.stm-assoc.org/industry-news/ssrn-launches/)

(本刊讯)